社会网络中关键节点的识别
——基于符号网络的 PageRank 算法改进

陈晓威　史昱天
(南京大学信息管理学院　南京 210023)

摘要:【目的】针对 PageRank 算法在符号网络中的局限性,提出其改进算法,以识别社会网络中的关键节点。【方法】基于符号网络的相关理论,将 PageRank 算法与点度中心性相结合,提出 KeyRank 算法,并对 Slashdot 网站的用户数据进行分析,以获取用户的 KeyRank 算法排名。【结果】PageRank 算法排名、入度排名、M-PR 算法排名与 KeyRank 算法排名在统计学意义上呈中度正相关。【局限】KeyRank 算法忽略了每次迭代时正、负链接的相互作用。【结论】传统算法与 KeyRank 算法在节点排序上存在差异,说明链接的符号属性对排序结果产生了重要影响,改进算法具有一定的理论和实践意义。

关键词: 符号网络　关键节点　PageRank 算法　点度中心性
分类号: TP301.6

1 引　言

随着互联网技术的迅速发展和社会化媒体的大规模普及,在线社交媒体的用户数量、整体规模得到显著提升,对社会化媒体的研究也从用户个体向用户之间相互关联所形成的网络拓扑结构延伸。

在众多有关社会网络的研究中,如何准确计算网络中节点的影响力一直是学者关注的重点,而在获取影响力的基础上,挖掘和识别其中的关键节点则是对社会网络更深层次的探索。然而在传统的社会网络分析中,研究者大多默认网络节点之间若存在关系,则均为正向关系,即支持、关注、合作等"示好"关系[1-3],因此,相关研究也普遍基于这种仅存在正向关系的网络结构识别关键节点,并未充分考虑边值的正负属性对研究结果的影响。随着社会网络研究层次的逐渐深入,一些研究者在对真实社会关系网络进行分析的过程中发现,节点之间除了存在正向、积极的友好关系外,还可能包含负向、消极的对立关系,其表达了一方对另一方的厌恶、反对等负面态度[4],这类同时包含"正向关系"(如顶、支持、肯定、点赞等)和"负向关系"(如踩、反对、否认、拉黑等)的社会网络结构,被统称为"符号网络"(Signed Network)[5]。符号网络提出至今,在人际关系、国际政治、信号传输等领域均有广泛应用,其节点间关系的正负属性能够更为客观地反映现实情况中的社会网络结构。在对节点影响力测算的过程中,将负向链接所传递的消极倾向纳入考虑范畴,分析节点负影响力产生的影响,更有利于掌握社会网络整体结构的动态演化,并提高识别关键节点的准确性与合理性。

因此,本文基于社会网络的链接特征,在综合考虑节点间正负向关联的基础上,提出一种改进的 PageRank 算法来识别网络中的关键节点。较之传统的识别算法,本文提出的 KeyRank 算法全面考察了网络链接的正负属性,并综合考虑节点的全局拓扑特性和局部特征,从而能够更为准确地识别出其中的关键节点,为基于符号网络的关键节点识别提供思路和帮助。

2 相关研究

关键节点识别作为社会网络分析的关键问题之一,自 20 世纪初期以来便被研究者所关注[6]。在信息

通讯作者:史昱天,ORCID:0000-0001-7347-4336,E-mail:njusyt@qq.com。

科学领域，随着对社会化媒体中信息传播和舆论扩散问题研究的深入，关于网络中关键节点的识别和挖掘也逐渐成为研究的热点。关键节点的位置及特征对社会网络中信息的传递路径、传播速度、覆盖范围等有着重要影响[7]，这些节点在网络结构动态演化、舆情传播控制等研究中都发挥着重要作用[8]。因此，如何快速准确地查找到网络中的关键节点，是研究社会网络拓扑结构和信息传播规律的必要环节。

社会网络中关键节点的识别方法可以分为两类：主观估测和客观度量。

(1) 主观估测是传统环境下关键节点识别的常用手段，主要包括自我报告法、知情人测量法和观察法[9]。在社会学和传播学领域，自我报告法是一种用于测度意见领袖影响力的有效方法，该方法基于设计的调查量表，让受访者根据实际情况作答，最后通过计算回答分值识别社会网络中的关键人物[10-11]；与自我报告法相对应的是知情人测量法，该方法通过询问他人而非自我识别网络中具有影响力的关键节点[12]；而观察法则通过观察研究对象在网络中的行为特征判定其重要性。经过 Rogers、King、Childers 及 Flynn 等学者的发展和完善，这些方法在社会科学各领域的研究中已被广泛接受和应用[13]。主观估测作为一种识别和评估关键节点的经典手段，却存在不适用于大规模网络和主观性较强等缺陷，因此，近年来涌现出大量以客观数据为基础的研究成果，以作为对关键节点识别方法的重要补充。

(2) 客观度量是基于网络节点及节点间关系等数据，采用社会网络分析、聚类分析、HITS 算法、PageRank 算法、数学建模等方法来识别关键节点。如陈远等以科学网博客社区为研究对象，运用社会网络分析法，分别构建博主好友关系网络和引文关系网络，分析网络的节点中心性和结构洞位置，得出关键节点与社会网络中位置的关系[14]；王珏等以国内某论坛为研究对象，引入 7 个特征值来度量意见领袖特性，提出基于 EM 算法的意见领袖识别算法，并通过聚类结果得到满足要求的子类[15]；熊涛等通过改进的 HITS 算法识别微博转发网络中的意见领袖，并发现粉丝数量与意见领袖的中心值高度正相关[16]；肖宇等提出基于无向、有权网络识别关键节点的 PageRank 算法，并以某论坛历史数据为例，验证其相比于其他算法具备更高的准确性[17]；Matumura 等提出的影响力扩散模型(Influence Diffusion Model)是利用数学建模识别关键节点的主要代表，其通过文本内容的相似度度量用户影响力，认为论坛中高影响力的节点为意见领袖[18]。

然而，上述关键节点识别的研究基础均为单一符号性质的网络环境，未对负向链接和缺失链接进行区分，与现实情境中节点间链接具有正负特性的情况并不相符，不能真实、客观地反映社会关系网络。因此，这类基于无符号网络的概念、算法及相关理论不能直接应用到同时包含正、负链接的符号网络中，研究者需要设计更具针对性的排序算法以挖掘符号网络中的关键节点[19]。

针对上述问题，已有学者基于符号网络环境开展了相关研究。识别符号网络中的关键节点可基于网络的链接结构，通过考虑正向链接和负向链接对网络中的一组节点进行排序[20]，排名较优的可认为是关键节点。因此，现有研究多从不同的角度考虑链接属性对节点的影响，从而对无符号网络中的关键节点识别算法进行改进。Bonacich 等基于社会网络分析法，提出一种度量符号网络特征向量中心性的方法，并以此分析修道士网络的派系结构及关键节点[21]；Li 等提出一个从用户生成内容中提取情感属性的框架，并以此构建符号网络，同时，通过聚类算法挖掘虚拟社区中的意见领袖[22]；Mishra 等基于 HITS 算法，考虑负向链接的影响，提出一种新的算法来度量信任网络中节点的偏好性与权威性[23]；Traag 等基于离散选择理论(Discrete Choice Theory)，考虑负向链接对节点排名的影响，同时借鉴 PageRank 算法的思想，提出“指数排名”算法，进而获得节点在网络中的全局可信度[24]。与国外研究相比，国内关于符号网络关键节点识别的研究还相对较少。顾洁等探究符号网络环境下节点影响力的测算方式，提出一种改进的 PageRank 算法，该算法将节点被正负链接指向的概率考虑在内，分别计算其在社交网络中的正、负影响力，以识别关键节点[25]。

综上所述，前人研究多以无符号网络环境作为关键节点识别的研究基础，基于符号网络环境的研究相对较少。因此，本文将 PageRank 算法这一重要的关键节点识别算法引入到符号网络中，通过将其与反映节点局部特征的点度中心性相结合，提出 KeyRank 算法，以挖掘符号网络中的关键节点。

3 基于符号网络的 PageRank 算法改进

Larry Page 和 Sergey Brin 于 1998 年提出的 PageRank 算法是评估网站页面重要性的经典算法，其认为网络环境下的网页链接结构和文献引文机制具有一定的相似性，从而将网页中的链接类比于文献之间的引用关系[26]。PageRank 算法的基本原理是"被越多高影响力的网页指向，则该网页的影响力越高"。例如，当网页 A 链接到网页 B 时就说明网页 A 为网页 B 投票，增加了网页 B 的影响力，即 PageRank 值，通过对整个网络的迭代计算，最终获得每个网页的 PageRank 值及相应的排名[27]。

3.1 基于符号网络的 PageRank 算法

PageRank 算法也是符号网络研究中最早使用的排序方法之一，其中，Shahriari 和 Jalili 将符号网络划分为 G^+(正向链接子图)和 G^-(负向链接子图)两个部分，并基于 PageRank 算法分别计算两个子图中各节点的 PageRank 值[28]，如公式(1)和公式(2)所示。

$$PR_i^+(t+1)=\alpha\sum_{j\in IN_i}\frac{PR_j^+(t)}{|Out_j^+|}+\frac{1-\alpha}{N} \tag{1}$$

$$PR_i^-(t+1)=\alpha\sum_{j\in IN_i}\frac{PR_j^-(t)}{|Out_j^-|}+\frac{1-\alpha}{N} \tag{2}$$

其中，IN_i 表示节点 i 的入链数，$|Out_j^+|$ 和 $|Out_j^-|$ 表示节点 j 的正向出链数和负向出链数，N 表示节点总数，α 表示阻尼系数，一般取值为 0.85。最终排序向量 PR 计算方法如公式(3)所示。

$$PR=PR^+-PR^- \tag{3}$$

虽然 PageRank 算法考虑到网络的全局拓扑特性，然而该算法忽略了网络中节点自身的属性[6,29]。从拓扑结构角度考虑，PageRank 算法弱化了局部属性对节点的影响。点度中心度被定义为网络中该节点的邻居节点的数目，用于反映一个节点在网络中的"权力"[1]，点度中心度越大说明节点在当前网络中的直接影响力越大，例如，被引次数较多的论文往往比较重要，微博中拥有较多粉丝数的用户影响力也往往较大。然而，PageRank 算法在计算节点的重要性时，虽然同时考虑了外部链接的数量和质量，但也在一定程度上削弱了邻居节点数量这一局部属性对节点的影响。图 1 提取自某社交网络，根据 PageRank 算法，若阻尼系数 $\alpha=0.85$，则 $PR(1)=0.2971$，$PR(2)=0.1853$，$PR(1)>PR(2)$，说明节点 1 比节点 2 重要；而根据节点入度的计算方法，$InDegree(1)=2$，$InDegree(2)=5$，$InDegree(1)<InDegree(2)$，说明节点 2 比节点 1 重要。可见，在该网络中，PageRank 算法在一定程度上弱化了入链数对节点的影响。

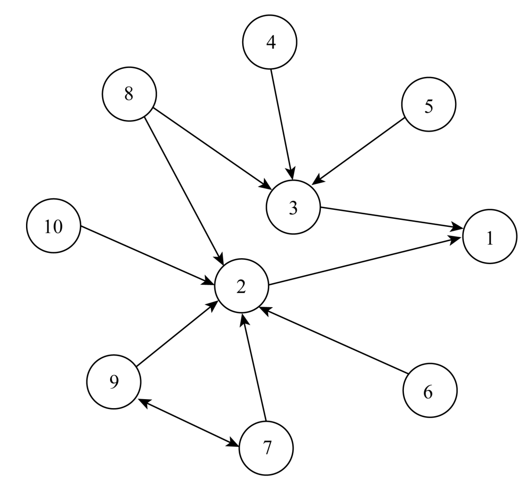


图 1 某社交网络图

在符号网络中，节点的局部特性对识别该节点是否为关键节点具有重要意义。从内容特征角度考虑，用户更喜欢与朋友的朋友而不是朋友的敌人建立正向链接；积极的用户更容易获取社会的支持，更可能被正向链接指向[30]；用户更容易同与他们相似的人建立正向链接[31]；低地位等级用户更可能与高地位等级用户建立正向链接[32]。而在社会网络情境中，朋友和积极的、相似的、高地位的用户在网络中一般可以认为是重要的、有影响的关键节点。因此，符号网络中节点的入链数和入链性质等局部特性同样是识别关键节点的重要因素。

3.2 基于 PageRank 的改进算法——KeyRank 算法

地位理论认为链接的符号决定了节点地位的差异，一条由 i 到 j 的正向链接 l_{ij}^+ 表示 i 认为 j 具有较高的地位，而负向链接 l_{ij}^- 表示 i 认为 j 具有较低的地位[33]。因此，一般认为正向链接入度有助于提升节点的地位，而负向链接入度则会降低节点的地位[32]。然而，在一些情景中，被负向链接指向的节点也可能成为关键节点，如识别网络谣言、虚假信息时，需要对网络节点负向影响力进行排名计算，便于管理者快速发现虚假信息，实现精确化辟谣[25]。

为识别社会网络环境中的关键节点，针对 PageRank

算法在符号网络中弱化了局部属性对节点的影响这一局限，笔者提出 KeyRank 算法。

(1) 基于 PageRank 算法思想，分别定义节点 i 在正向链接网络和负向链接网络中的 PageRank 值 PR_i^+ 和 PR_i^-，计算方法如公式(4)和公式(5)所示。

$$PR_i^+ = \alpha \sum_{j \in IN_i^+} \frac{PR_j^+}{|Out_j^+|} + \frac{1-\alpha}{N} \quad (4)$$

$$PR_i^- = \alpha \sum_{j \in IN_i^-} \frac{PR_j^-}{|Out_j^-|} + \frac{1-\alpha}{N} \quad (5)$$

其中，IN_i^+ 和 IN_i^- 表示节点 i 的正向入链数和负向入链数。

(2) 在 PR_i^+ 和 PR_i^- 基础上，将其与反映网络局部特征的点度中心度结合，考虑入度对节点的影响，从而定义节点 i 在正向链接网络和负向链接网络中的影响力值 PP_i 和 NP_i，计算方法如公式(6)和公式(7)所示。

$$PP_i = (PR_i^+)^{\beta} \cdot |In_i^+|^{1-\beta} \quad (6)$$

$$NP_i = (PR_i^-)^{\beta} \cdot |In_i^-|^{1-\beta} \quad (7)$$

其中，$|In_i^+|$ 和 $|In_i^-|$ 表示节点 i 的正向链接入度和负向链接入度，参数 $\beta \in [0,1]$，用于调节不同情境下 PageRank 值与入度对关键节点识别影响的比例。当 $\beta = 1$ 时，PP_i 和 NP_i 退化为 PageRank 值；当 $\beta = 0$ 时，PP_i 和 NP_i 退化为节点的入度。

(3) 一般地，定义节点 i 的重要度 $KeyRank_i$，用于识别关键节点，计算方法如公式(8)所示。

$$KeyRank_i = PP_i - NP_i \quad (8)$$

$KeyRank_i$ 越大，表示节点越重要、影响力越大、地位越高。

但不是所有的情景都认为负向链接不重要，如识别争议人物、负面信息时，负向链接同正向链接一样重要，此时，需重新定义 $KeyRank_i$，同时定义节点 i 的情感倾向 $EmotionRank_i$，计算方法如公式(9)和公式(10)所示。

$$KeyRank_i = PP_i + NP_i \quad (9)$$

$$EmotionRank_i = PP_i - NP_i \quad (10)$$

$KeyRank_i$ 越大，表示节点越重要、影响力越大、地位越高；$EmotionRank_i > 0$，表示节点 i 的情感为正，该值越大说明节点 i 的正向情感越强烈；$EmotionRank_i < 0$，表示节点 i 的情感为负，该值越小说明节点 i 的负向情感越强烈。

4　实证研究

4.1　实验数据

国外学者针对符号网络的研究数据主要来自于商品评论网站 Epinions、科技资讯评论网站 Slashdot 以及协同编辑在线百科全书 Wikipedia 等[5, 28]。本文选取 Slashdot 网站的社交数据，该网站提供的 Slashdot Zoo 功能可以让用户将他人标记为“朋友”或者“敌人”，从而构建了同时包含正向链接与负向链接的符号网络结构。实验数据来自斯坦福网络数据库(SNAP)，所选取的数据集包含了 Slashdot 网站截至 2009 年 2 月 21 日的社交网络数据，共有 82 144 个节点和 549 202 组关联信息[5]。

4.2　基于 KeyRank 算法的实验结果

基于 KeyRank 算法，根据公式(4)–公式(8)，令 $\alpha = 0.85$，分别计算 β 在不同取值时各节点的 $KeyRank$ 值，并进行排名。表 1 列出了在 $\beta = 0$、$\beta = 0.25$、$\beta = 0.5$、$\beta = 0.75$ 和 $\beta = 1$ 时，Slashdot 数据集中排在前 20 位的节点。其中，当 $\beta = 0$ 时，结果即为基于符号网络的节点入度排名；当 $\beta = 1$ 时，结果即为基于符号网络的 PageRank 值排名。

表 1　各节点的 KeyRank 排名结果(前 20 名)

Rank	$\beta = 0$	$\beta = 0.25$	$\beta = 0.5$	$\beta = 0.75$	$\beta = 1$
1	937	937	937	90	90
2	1 935	1 935	90	937	937
3	90	90	1 935	1 935	1 935
4	531	531	1 485	1 485	1 485
5	1 485	1 485	531	1 930	1 930
6	1 930 433	1 930	1 930	531	1 635
7	—	2 208	2 208	1 635	198
8	2 208	1 635	1 635	2 208	2 208
9	179	179	179	59	531
10	5 128	59	59	179	59
11	928	7 821	7 821	198	179
12	1 802	1 802	1 802	7 821	7 821
13	7 821	433	928	1 802	1 050
14	1 635	928	686	1 050	1 802
15	686	686	2 023	176	176
16	534	5 363	678	686	686
17	5 363	2 023	433	928	2 023
18	59	678	5 363	2 023	928
19	9 835	5 128	176	678	678
20	10 762	534	198	5 363	1 953

为探究 β 在不同取值时各排名之间的关系，对排名结果进行 Kendall's tau-b 相关系数检验。由于网络中存在大量入度为 0 的节点，为避免其对分析结果造成干扰，故先将这些数据剔除后再进行分析，结果如表 2 所示。根据各排名结果的相关系数矩阵可知，当 β 取值越相近时，其排名结果越一致；当 $\beta=0$ 与 $\beta=1$ 时，其排名结果的相关系数最低，$\tau=0.5857$，说明基于符号网络的节点入度排名和 PageRank 值排名在统计学意义上呈中度正相关，两种算法存在一定差异。

表 2　各排名结果的相关系数矩阵

β 值	0	0.25	0.5	0.75	1
0	1	0.8763	0.7746	0.6486	0.5857
0.25	0.8763	1	0.9070	0.7918	0.7050
0.5	0.7746	0.9070	1	0.8847	0.7459
0.75	0.6486	0.7918	0.8847	1	0.7640
1	0.5857	0.7050	0.7459	0.7640	1

(*显著性 p 值均远小于 0.05。)

4.3　传统算法与 KeyRank 算法的比较

为比较传统算法与 KeyRank 算法在排名结果上的差异，本文对无符号网络中各节点的 PageRank 值和入度进行分析，同时，将 Shahriari 和 Jalili 所提出的改进算法(Modified PageRank, M-PR)作为比较对象。

在传统研究中，研究者更加关注网络中那些正向、积极的链接，因此，在计算节点影响力时仅考虑了符号网络中正向链接子图 G^+的影响，不对负向链接和缺失链接进行区分[31]。以信息科学领域为例，引文网络中的合作关系、引用关系，社交网络中的朋友关系、支持关系，均为正向链接，因此，可提取网络中的正向链接子图，以实现符号网络的无符号化。此外，为实现符号网络的无符号化，也有学者将符号网络中的正向链接和负向链接视为同质的链接[34]。以信息科学领域为例，在线投票网站的投票者之间意见无论相同或相左，电商网站中用户对商品的评论无论好评或差评，都视作存在关系。综上所述，本文分别采用以上两种方式对 Slashdot 网络中的链接进行无符号化处理。

将剔除了负向链接的 Slashdot 网络记为 Network A，将正向链接和负向链接同质化的 Slashdot 网络记为 Network B；同样地，剔除 Network A 和 Network B 中入度为 0 的节点；最后对 PageRank 算法、入度、M-PR 算法和 KeyRank 算法($\beta=0.5$)的排名结果进行 Kendall's tau-b 相关系数检验，结果如表 3 和表 4 所示。可知，PageRank 算法排名、入度排名、M-PR 算法排名与 KeyRank 算法排名在统计学意义上呈中度正相关，说明传统算法与 KeyRank 算法在节点排序上存在差异；此外，M-PR 算法与 KeyRank 算法在排序结果上的相关系数较高，说明考虑了网络正、负链接影响的算法在排序结果上具有一致性。

表 3　基于 Network A 的相关系数矩阵

	PageRank	Indegree	M-PR	KeyRank
PageRank	1	0.5545	0.7607	0.5417
Indegree	0.5545	1	0.3562	0.4547
M-PR	0.7607	0.3562	1	0.7022
KeyRank	0.5417	0.4547	0.7022	1

(*显著性 p 值均远小于 0.05。)

表 4　基于 Network B 的相关系数矩阵

	PageRank	Indegree	M-PR	KeyRank
PageRank	1	0.5709	0.5659	0.4401
Indegree	0.5709	1	0.2706	0.3234
M-PR	0.5659	0.2706	1	0.7459
KeyRank	0.4401	0.3234	0.7459	1

(*显著性 p 值均远小于 0.05。)

5　结　语

社会网络中关键节点的识别一直是信息科学领域研究的重要议题，然而，学者多基于无符号网络进行研究，忽略了网络中链接的正负属性。传统基于无符号网络设计的算法在符号网络中已不再适用，因此，本文基于符号网络的相关理论，从网络拓扑特征出发，提出一种改进的 PageRank 算法——KeyRank 算法，用于识别符号网络中的关键节点。

改进算法在 PageRank 算法的基础上，结合基于中心性的统计方法，综合考虑节点的全局拓扑特性和局部特征，并设置参数 β 来调节不同情境下"全局"与"局部"对关键节点识别影响的比重。在实证研究中发现，传统算法与 KeyRank 算法在节点排序上存在差异，说明链接的符号属性对排序结果产生了重要影响。基于符号网络设计的 KeyRank 算法与传统算法的区别主要体现在两方面：一是更加关注节点自身的属性特征，通过考虑入链数和入链性质识别关键节点；二是对不同情境下节点的重要程度进行区分，一般地，负向链

接对识别关键节点只产生负面影响，但有时负向链接却对关键节点的识别具有重要意义。

在本研究的基础上，未来可以重点考虑如下工作：

(1) 改进 KeyRank 算法，该算法仅在计算节点重要度时考虑了正负链接对节点的影响，忽略了每次迭代时正负链接的相互作用。

(2) 通过实证研究，给定不同情境下参数 β 的经验取值。β 的取值视不同的网络结构和背景信息而定，但相关情景和应用场合的 β 值相近。

(3) 在拓扑结构特征的基础上，加入对符号网络中内容特征和行为特征的考量，识别关键节点。拥有背景信息的符号网络数据还较难获取，相关研究在一定程度上受到限制。

(4) 符号网络作为近年来新兴的研究热点，理论基础、研究方法和实证环境尚不成熟。因此，基于符号网络的排序、分类、聚类、预测、信息传播和推荐等问题还有较大的研究和发展空间。

参考文献：

[1] 朱庆华，李亮. 社会网络分析法及其在情报学中的应用[J]. 情报理论与实践，2008，31(2)：179-183. (Zhu Qinghua, Li Liang. Social Network Analysis Method & Its Application in Information Science [J]. Information Studies: Theory & Application, 2008, 31(2): 179-183.)

[2] 王曰芬，杭伟梁，丁洁. 微博舆情社会网络关键节点识别与应用研究[J]. 情报资料工作，2016，37(3)：6-11. (Wang Yuefen, Hang Weiliang, Ding Jie. Identification and Application of Microblog Public Opinion Social Network Critical Node[J]. Information and Documentation Services, 2016, 37(3): 6-11.)

[3] 王陆. 虚拟学习社区的社会网络分析[J]. 中国电化教育，2009(2)：5-11. (Wang Lu. Social Network Analysis in Virtual Learning Community [J]. China Educational Technology, 2009(2): 5-11.)

[4] 程苏琦，沈华伟，张国清，等. 符号网络研究综述[J]. 软件学报，2014，25(1)：1-15. (Cheng Suqi, Shen Huawei, Zhang Guoqing, et al. Survey of Signed Network Research[J]. Journal of Software, 2014, 25(1): 1-15.)

[5] Leskovec J, Huttenlocher D, Kleinberg J. Signed Networks in Social Media[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010: 1361-1370.

[6] 韩忠明，陈炎，刘雯，等. 社会网络节点影响力分析研究[J]. 软件学报，2017，28(1)：84-104. (Han Zhongming, Chen Yan, Liu Wen, et al. Research on Node Influence Analysis in Social Networks [J]. Journal of Software, 2017, 28(1): 84-104.)

[7] 康伟. 基于 SNA 的突发事件网络舆情关键节点识别——以“7·23 动车事故”为例[J]. 公共管理学报，2012，9(3)：101-111. (Kang Wei. Analysis of the Key Nodes in Public Opinion Spread During Emergencies Based on Social Network Theory——A Case Study of the 7·23 Wenzhou High-speed Train Collision [J]. Journal of Public Management, 2012, 9(3): 101-111.)

[8] Zhou T, Wang B H. Catastrophes in Scale-free Networks[J]. Chinese Physics Letters, 2005, 22(5): 1072-1075.

[9] 陈华珊. 业主论坛意见领袖——识别方法及其特点[J]. 青年研究，2013(6)：65-72. (Chen Huashan. Measurement, Identification and Characteristics of Opinion Leaders in Real Estate Owner Online Forum [J]. Youth Studies, 2013(6): 65-72.)

[10] King C W, Summers J O. Overlap of Opinion Leadership across Consumer Product Categories[J]. Journal of Marketing Research, 1970, 7(1): 43-50.

[11] Flynn L R, Goldsmith R E, Eastman J K. Opinion Leaders and Opinion Seekers: Two New Measurement Scales[J]. Journal of the Academy of Marketing Science, 1996, 24(2): 137-147.

[12] 梦非. 社会化商务环境下意见领袖对购买意愿的影响研究[D]. 南京：南京大学，2012. (Meng Fei. Research of Opinion Leaders’ Impact on Purchase Intention Under Social Commerce Context[D]. Nanjing: Nanjing University, 2012.)

[13] Dasgupta S, Momtaz N J, Aghaie A, et al. Identifying Opinion Leaders for Marketing by Analyzing Online Social Networks[J]. International Journal of Virtual Communities & Social Networking, 2011, 3(1): 43-59.

[14] 陈远，刘欣宇. 基于社会网络分析的意见领袖识别研究[J]. 情报科学，2015，33(4)：13-19，92. (Chen Yuan, Liu Xinyu. Research on Opinion Leaders Recognition Based on Social Network[J]. Information Science, 2015, 33 (4): 13-19, 92.)

[15] 王珏，曾剑平，周葆华，等. 基于聚类分析的网络论坛意见领袖发现方法[J]. 计算机工程，2011，37(5)：44-46. (Wang Jue, Zeng Jianping, Zhou Baohua, et al. Online Forum Opinion Leaders Discovering Method Based on Clustering Analysis[J]. Computer Engineering, 2011, 37(5): 44-46.)

[16] 熊涛，何跃. 微博转发网络中意见领袖的识别与分析[J]. 现代图书情报技术，2013(6)：55-62. (Xiong Tao, He Yue. The Identification and Analysis of Micro-blogging Opinion Leaders in the Network of Retweet Relationship[J]. New Technology of Library and Information Service, 2013(6):

55-62.)

[17] 肖宇, 许炜, 夏霖. 网络社区中的意见领袖特征分析[J]. 计算机工程与科学, 2011, 33(1): 150-156. (Xiao Yu, Xu Wei, Xia Lin. A Feature Analysis of the Opinion Leader in On-Line Communities[J]. Computer Engineering and Science, 2011, 33(1): 150-156.)

[18] Matsumura N, Ohsawa Y, Ishizuka M. Mining and Characterizing Opinion Leaders from Threaded Online Discussions[C]//Proceedings of the 6th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies. 2002: 1267-1270.

[19] Chiang K Y, Hsieh C J, Natarajan N, et al. Prediction and Clustering in Signed Networks: A Local to Global Perspective[J]. The Journal of Machine Learning Research, 2014, 15(1): 1177-1213.

[20] Getoor L, Diehl C P. Link Mining: A Survey[J]. ACM SIGKDD Explorations Newsletter, 2005, 7(2): 3-12.

[21] Bonacich P, Lloyd P. Calculating Status with Negative Relations[J]. Social Networks, 2004, 26(4): 331-338.

[22] Li X, Chen H, Li S. Exploiting Emotions in Social Interactions to Detect Online Social Communities[C]// Proceedings of Pacific Asia Conference on Information Systems, PACIS 2010, Taipei, Taiwan, China. 2010.

[23] Mishra A, Bhattacharya A. Finding the Bias and Prestige of Nodes in Networks Based on Trust Scores[C]//Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India. 2011: 567-576.

[24] Traag V A, Nesterov Y E, Van Dooren P. Exponential Ranking: Taking into Account Negative Links [C] //Proceedings of the International Conference on Social Informatics.Springer Berlin Heidelberg, 2010.

[25] 顾洁, 胡安安, 刘旭, 等. 社交网络正、负影响力计算——基于符号网络的 PageRank 算法改进[J]. 情报学报, 2015, 34(7): 725-733. (Gu Jie, Hu An'an, Liu Xu, et al. Analysis of Positive and Negative Influential Power in Social Networks——Improving PageRank in Signed Networks[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(7): 725-733.)

[26] Page L, Brin S, Motwani R, et al. The PageRank Citation Ranking: Bringing Order to the Web[R]. Stanford InfoLab, 1999.

[27] 曹军. Google 的 PageRank 技术剖析[J]. 情报杂志, 2002, 21(10): 15-18. (Cao Jun. The Anatomy of Page Rank Technology of Google[J]. Journal of Intelligence, 2002, 21(10): 15-18.)

[28] Shahriari M, Jalili M. Ranking Nodes in Signed Social Networks[J]. Social Network Analysis and Mining, 2014, 4(1): 1-12.

[29] 刘建国, 任卓明, 郭强, 等. 复杂网络中节点重要性排序的研究进展[J]. 物理学报, 2013, 62(17): 178901. (Liu Jianguo, Ren Zhuoming, Guo Qiang, et al. Node Importance Ranking of Complex Networks[J]. Acta Physica Sinica, 2013, 62(17): 178901.)

[30] Beigi G, Tang J, Liu H. Signed Link Analysis in Social Media Networks[C]//Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM 2016), Cologne, Germany. 2016: 539-542.

[31] Tang J, Chang Y, Aggarwal C, et al. A Survey of Signed Network Mining in Social Media [J]. ACM Computing Surveys, 2014, 9(4): Article 39. DOI: 10.1145/0000000.0000000.

[32] 王鑫, 王英, 左万利. 基于交互意见和地位理论的符号网络链接预测模型[J]. 计算机研究与发展, 2016, 53(4): 764-775. (Wang Xin, Wang Ying, Zuo Wanli. Exploring Interactional Opinions and Status Theory for Predicting Links in Signed Network[J]. Journal of Computer Research and Development, 2016, 53(4): 764-775.)

[33] Zheng X, Zeng D, Wang F Y. Social Balance in Signed Networks[J]. Information Systems Frontiers, 2015, 17(5): 1077-1095.

[34] Kunegis J, Lommatzsch A, Bauckhage C. The Slashdot Zoo: Mining a Social Network with Negative Edges [C]// Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain. ACM, 2009: 741-750.

作者贡献声明:

陈晓威: 提出研究思路, 设计研究方案, 算法设计, 进行实验, 论文撰写;

史昱天: 采集、清洗和分析数据, 论文修改及最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: njusyt@qq.com。

[1] 史昱天. Slashdot.csv. Slashdot 网站用户关系数据集.

收稿日期: 2017-06-12
收修改稿日期: 2017-07-26

Identifying Key Nodes in Social Network with Improved PageRank Algorithm

Chen Xiaowei　Shi Yutian
(School of Information Management, Nanjing University, Nanjing 210023, China)

Abstract: [**Objective**] This paper modifies the PageRank algorithm for signed network, aiming to identify the key nodes in social network. [**Methods**] Based on the theory of signed network, we proposed the KeyRank algorithm, which combined the PageRank algorithm with node centrality. We examined the new algorithm with user data from the Slashdot website to obtain every user's ranking. [**Results**] The rankings of PageRank algorithm, in-degree and M-PR algorithm had significant medium level positive correlation with the rankings obtained with the KeyRank algorithm. [**Limitations**] The KeyRank algorithm ignored the interactions between the positive and negative links in each iteration. [**Conclusions**] There is difference between the rankings of nodes by traditional and KeyRank algorithms. The signed links poses important impacts on the rankings, which shows the improved algorithm's theoretical and practical significance.

Keywords: Signed Network　Key Nodes　PageRank Algorithm　Node Centrality